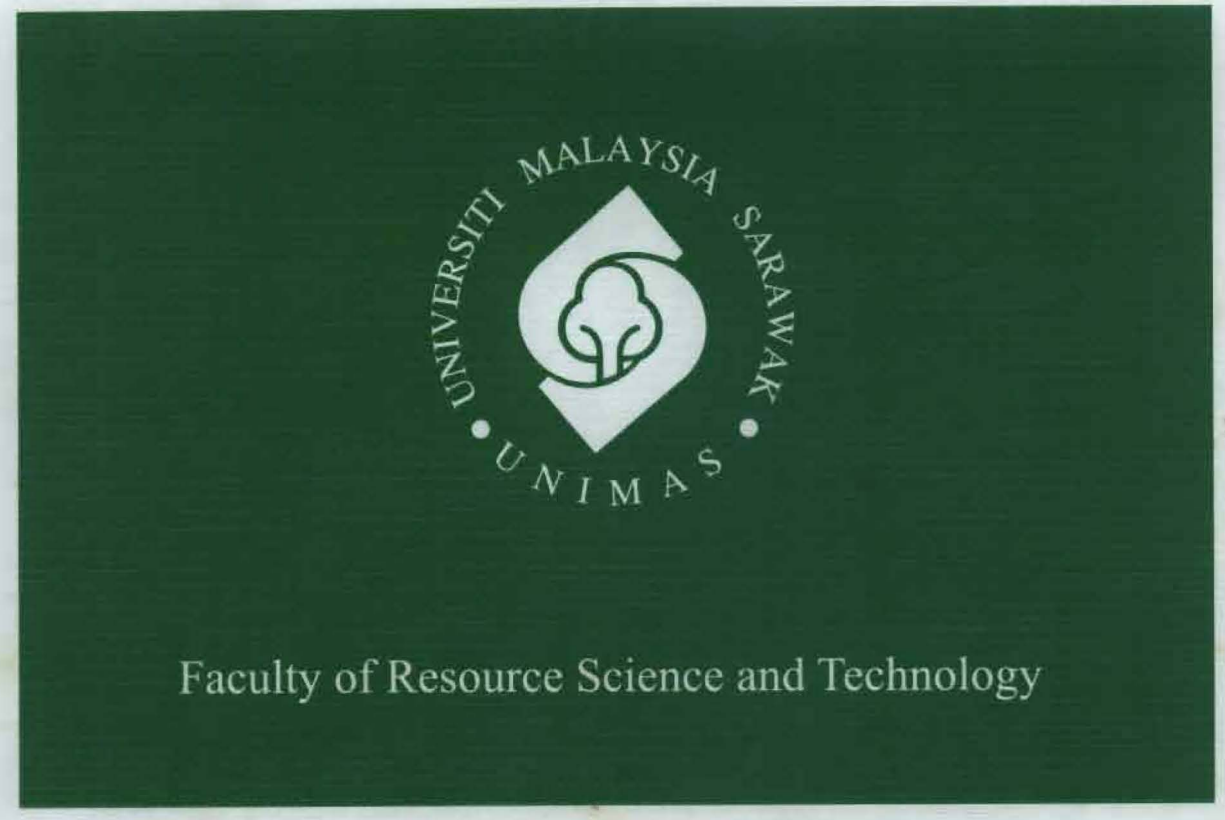

OPTIMIZATION OF DNA EXTRACTION FROM ORANG UTAN USING MITOCHONDRIAL CYTOCHROME *b* GENE

Kennedy Lim Chong Ming (21194)

Bachelor of Science with Honours
(Resource Biotechnology)
2011

OPTIMIZATION OF DNA EXTRACTION FROM ORANG UTAN USING MITOCHONDRIAL CYTOCHROME *b* GENE

KENNEDY LIM CHONG MING

This project is submitted in partial fulfillment of the requirements for the degree of
Bachelor of Science with Honours
(Biotechnology)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK
2011

ACKNOWLEDGEMENT

I would like to express my deepest gratitude and sincere thanks to all who have contributed for the completion of this thesis. This includes my supervisor, Dr. Yuzine Esa, a person that I much respect and honoured to be under his supervision. His continuous advice, guidance and motivation had helped me throughout this research and in this completion of thesis.

Likewise, I would also like to thank to all the officers in Matang Wildlife Centre (MWC) for lending their help in providing samples of the Orang Utans and allowing me to conduct research on it. I would also like to expand my sincere gratitude to all of the postgraduate students of zoology department for their help in giving advice, suggestion, encouragement and support for my research study.

Lastly, I would like to thank my family for their devotion in believing me and for unforgettable emotion support. Not forgetting to all my colleagues and friends who have helped me intentionally or unintentionally, physically or emotionally, in completing this research. This thesis would not be completed without the help of the people mentioned above.

Declaration

I hereby declare that this thesis entitled "Optimization of DNA Extraction from Orang Utan Using Mitochondrial Cytochrome *b* Gene" is the result of my own research work and effort. It has not been submitted anywhere for any award. Where other sources of information have been used, they have acknowledged.

Signature:

Name: KENNEDY LIM CHONG MING

Date: 25/5/2011

Table of Contents

List of Abbreviations

bp - base pair

CTAB - Cetyltrimethyl Ammonium Bromide

DNA - Deoxyribonucleic acid

dNTP - Deoxyribonucleotide triphosphate mixtures

EDTA - Ethlenediaminetetraacetic acid

EtBr - Ethidium Bromide

IUCN - International Union for Conservation of Nature

MEGA - Molecular Evolutionary Genetic Analysis

$MgCl_2$ - Magnesium chloride

mtDNA - Mitochondrial DNA

NaCl - Natrium Chloride

P. pygmaeus - *Pongo pygmaeus*

P. p. pygmaeus - *Pongo pygmaeus pygmaeus*

P. p. morio - *Pongo pygmaeus morio*

P. p. wurmbii - *Pongo pygmaeus wurmbii*

PCR - Polymerase Chain Reaction

List of Tables and Figures

Tables

Figures

Optimization of DNA Extraction from Orang Utan Using Mitochondrial Cytochrome *b* Gene

Kennedy Lim Chong Ming

Resource Biotechnology Programme
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSRACT

This study examines the optimization of DNA extraction from Orang Utans' hair samples. The DNA sequence analyses were carried out using mitochondrial Cytochrome *b* gene. A total of 8 samples of Orang Utan were used in this study, which were collected from the Matang Wildlife Centre. The methods used were conventional CTAB method and commercial DNA Extraction Kit (Promega). The results showed that extraction was possible using both methods but the DNAs present were contaminated with organic materials. Due to the contaminants and insufficient DNA concentration, optimization of the extracted DNA was not successfully achieved.

Key words: Optimization, DNA extraction, Orang Utan, mitochondrial cytochrome *b* gene, conventional CTAB method.

ABSTRAK

Kajian ini bertujuan untuk mengoptimumkan pengekstrakan DNA yang dari sampel rambut Orang Utan. Analisis jujukan DNA dijalankan menggunakan gen mitokondria sitokrom b. Sejumlah 8 sampel Orang Utan telah digunakan dalam kajian ini, yang dikutip dari Pusat Hidupan Liar Matang. Kaedah-kaedah yang digunakan adalah kaedah konvensional CTAB dan komersial Kit Pengekstrakan DNA (Promega). Keputusan menunjukkan kedua-dua kaedah itu boleh digunakan untuk mengekstrak DNA tetapi terdapat juga kehadiran bahan-bahan organik. Disebabkan kehadiran bahan organik dan kepekatan DNA yang kurang, pengoptimuman DNA yang diekstrak tidak berjaya dicapai.

Kata kunci: Pengoptimuman, pengekstrakan DNA, Orang Utan, gen mitokondria sitokrom b, kaedah konvensional CTAB.

1.0 Introduction

Genetic (DNA-based) studies, DNA profiling and other methods have been widely used in research and it is one of the major contributors in gathering information on the diversity, conservation biology and population genetic analysis of the endangered or threatened species (Avise, 1996; Snow and Parker, 1998).

One of the contributions of these technologies of genetic-based is gathering information of the life-threatened *Pongo pygmaeus*. According to Warren *et al.* (2001), it is known that within the Bornean population, three subspecies of *Pongo pygmaeus* are identified; *Pongo pygmaeus pygmaeus* in Southwest and Central Kalimantan, *Pongo pygmaeus wurmbii* in Sarawak and *Pongo pygmaeus morio* in Sabah and East Kalimantan. The study suggested that the Bornean Orang Utans haven't faced a serious of clustering in its population. Warren *et al.* (2001) also predicted that the Bornean and Sumatran Orang Utans diverged around 1.1 million years ago and the subpopulation diverged around 860,000 years ago.

Mitochondrial DNA (mtDNA) has been used as genetic marker as it has simple sequence organization, maternal inheritance and absence of recombination. Thus it makes an ideal marker for tracing maternal genealogies. According to Ballard (2000), the evolution of the mtDNA seems not to be affected by recombination. In the absence of recombination, the sequences maintain the phylogenetic signal and all the differences among the sequence are due to mutations. Sequence analysis of the most variable segment of the control region of the rapidly evolving mtDNA molecule has long been the method of choice for analysing the population level diversity in human and in great apes (Vigilant *et al.* 1991; Morin *et al.* 1993).

Extraction is crucial as this can lead to many varied applications. Through extraction, DNA analyses are possible to be made and further investigation can be applied as well. Through non-invasive method of sampling ensure that the endangered or threatened species are well kept, conserved and not harmed. In this project, hair samples were taken and subjected to extraction for DNA. Optimization for the hair sampling is somewhat difficult to do compare to other samples such as blood, liver tissue and so on.

Thus, the objectives of this study are:

1) To extract DNA from Orang Utans' hair samples, and
2) To amplify mitochondrial cytochrome *b* gene from the extracted DNA.

2.0 Literature Reviews

2.1 Orang Utan (*Pongo pygmaeus* in Borneo, *Pongo abelii* in Sumatra)

The name "Orang Utan" comes from a local language, where in Malay "orang" means "person" and the word "utan" derived from the word "hutan", which means "forest". Thus, it literally translated as "person of the forest". Million years ago, the Orang Utan lives throughout Asia. But today they can only be found in the island of Borneo and Sumatra. According to Warren *et al.* (2001), the Bornean Orang Utans are divided into 3 more subspecies at particular region and geographic clustering:

i. *Pongo pygmaeus pygmaeus* can be found in Northwest of Borneo (Sarawak, Malaysia and Northwest Kalimantan, Indonesia), where Figure 1 shows one of the Orang Utan available in Matang Wildlife Centre, Sarawak,

ii. *Pongo pygmaeus wurmbii* in Central Borneo (Southern West Kalimantan and Central Kalimantan, Indonesia) and;

iii. *Pongo pygmaeus morio* can be found in Northeast Borneo (Sabah, Malaysia and East Kalimantan, Indonesia).

The two species, *Pongo abelii* (Sumatra) and *Pongo pygmaeus* (Borneo), shows a slightly different physical characteristic. The Sumatran Orang Utans have a narrower face and longer beard compared to the Bornean species. The Bornean Orang Utans are slightly darker in colour and the males have wider cheek pads. Their behavioural and diet are also been observed, where the Sumatran Orang Utans are more frugivorous (fruit-eating) and there is evidence of tool use, compared to Bornean Orang Utans. Under the International Union for Conservation of Nature (IUCN) Red List, where the distribution of their location can be seen and shown on Figure 2, the Sumatran Orang Utan is classified as critically endangered and the Bornean Orang Utan as endangered species.

Figure 1: Orang Utan in Matang Wildlife Centre, Sarawak (http://www.orangutanproject.com/content.php?p=59)

- Class : Mammalia
- Order : Primates
- Super family : Hominoidea
- Family : Pongidae
- Genus : *Pongo*
- Species : *P. abelii* (Sumatran) and *P. pygmaeus* (Bornean)
- Length : males – about 40 inches from top of head to rump;
  females – about 30 inches
- Weight : males – 110 to 300 pounds;
  females – 66 to 110 pounds
- Life Span : 60 years or more
- Age of Maturity : males – about 15 years;
  females – about 12 years (in captivity)
- Conservation Status : *Pongo pygmaeus* (Bornean) is endangered;
  Pongo abelii (Sumatran) is critically endangered

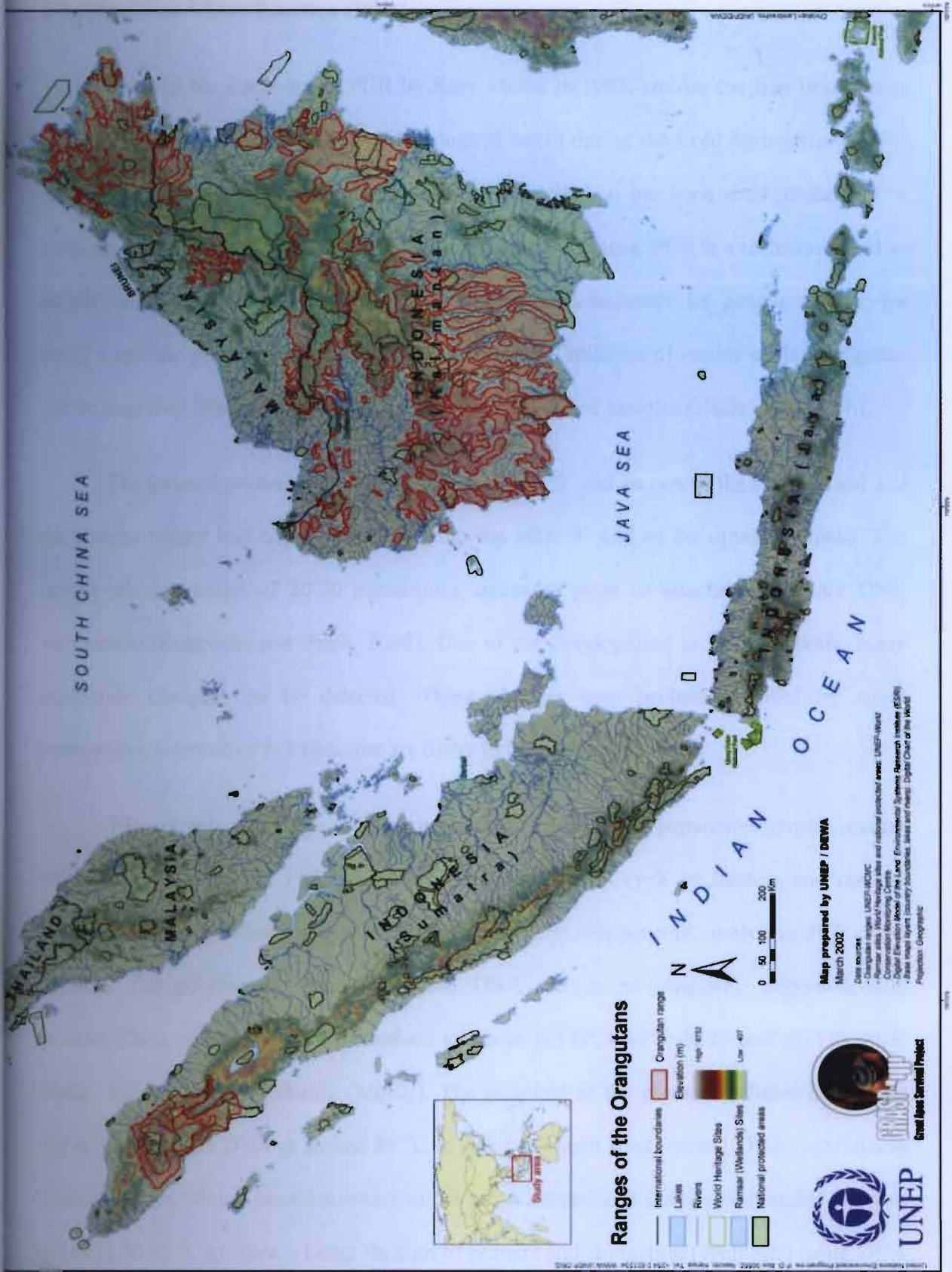

Figure 2: Range of distribution of Orang Utan. Source: UNEP (http://www.unep.org/grasp/docs/OrangUtan.pdf)

2.2 Polymerase Chain Reaction (PCR)

Through the discovery of PCR by Kary Mullis in 1985, and for the first time it was introduced formerly to the medical and biological world during the Cold Spring Harbor 51st Symposium on Quantitative Biology (Mullis *et al.*, 1986), it has been used intensively in research, medical and all that associated with biological sciences. PCR is a technique used to amplify a fragment or a specific region of desired DNA sequence for genetic studies. By using a specific primers (forward and reverse primers), millions of copies of desired genes can be amplified from just a small amount of tissues or blood samples (Hillis *et al.*, 1996).

The forward primers will complementary to the 3' end on one of the DNA strand and the reverse primer will be complementary on the other 3' end of the opposite strand. The primer which consists of 20-30 nucleotides serves as point of attachment for the DNA polymerase (Kolmodin and Birch, 2002). Due to the development in PCR methods, many nucleotide changes can be detected. These changes may probably caused by silent substitution, insertion or deletion, that are likely to be selectively natural.

Through this advantage, PCR is a suitable technique for population genetic studies (Silva and Russo, 2000). PCR works by a repetitive series cycle of heating and cooling (Henry, 1992) that defined as one PCR cycle. Common components involve in PCR cycle includes Taq polymerase, double stranded DNA (serves as template), oligonucleotide primers, Deoxyribonucleotide triphosphate mixtures (dTTP, dATP, dCTP and dGTP), PCR buffer, and Magnesium chloride ($MgCl_2$). The principal of the process includes denaturing the double stranded DNA at around 95 °C in order to obtain single-strand DNA, hybridizing (annealing) the primers complementary to the target sequence at the single-stranded DNA in between 40-65 °C to allow a better fixation of primers and elongation (synthesis) using DNA polymerase (*Taq* polymerase) starting from the primers at 72 °C to make a new copy of the

single strand DNA. This process is repeated until the desired amounts of the target sequence are enough to continue to the next step.

2.3 DNA sequencing

DNA sequencing is a term used for comparing population in high resolution and facilitation interpretation, by determining the order of the constituent bases (adenine, guanine, cytosine, and thymine) of deoxyribonucleic acid (DNA) (Hoelzel and Dover, 1991). It is a powerful tool for measurement of genetic characterization among population and species (Henry, 1992), thus the best method in soughting out detection genetic variation and phylogenetic relationship between species and population.

The evolution in technologies provides researches a more convenient way and speed up biological research and discovery of new things. This is achieved by the evolutionary if DNA sequencing that help in constituting and constructing phylogenetic trees to understand more on genetic relation among species.

2.4 Mitochondrial DNA

Mitochondrial DNA (mtDNA) contains 37 genes, where thirteen of these genes involve in giving orders in making enzyme involved in oxidative phosphorylation process. Through this process, energy is created in the form of adenosine triphosphate (ATP), which is the main source of energy for cell. The remaining genes codes for transfer RNA (tRNA) and ribosomal RNA (rRNA). Different form nuclear DNA which is inherited equally from both mother and father, mtDNA is inherited only from the mother, as all the mitochondria are inherit from the maternal's egg cells, and thus useful for tracing individuals' maternal lineage

(Hoelzel and Dover, 1991). Sequence analysis using mtDNA molecule has been established as the method of choice for analysing population level diversity in humans and great apes (Vigilant *et al.* 1991; Morin *et al.* 1993).

Thus, mtDNA has the advantage of being a marker for studying genetic variation. MtDNA is beneficial to forensic scientists as (Lotter, 2008):

i. Samples that is old or has been degraded can be tested as the structural and location of the mtDNA in the cell are stable as it is buried deep within the cell and has a circular structure which protects it from deterioration,

ii. It can be found in a large quantities per cell, thus small sample can be tested, and;

iii. It can be extracted from samples such as hair shafts and bone fragments whereas nuclear DNA cannot.

Since it can be amplified easily and because it is haploid, the sequence can be obtained without cloning. The mtDNA has a high evolutionary rate and this opens a chance for it to recover the pattern and tempo of recent historical events without giving an extensive sequencing effort. As it has at least low recombination area, the whole molecule can be considered to have the same genealogical history. Through these reasons, mtDNA has been widely used as marker for phylogenetic studies.

2.5 Cytochrome *b*

Cytochrome *b* (cyt *b*) has been widely used as a primer to amplify a specific region of the DNA. Considering as a valuable gene to identify the phylogeny of the species, cytochrome b is well known mitochondrial gene with the respect of its structural and functional of its protein product (Esposti *et al.,* 1993), and its involvement in electron transport chain of mitochondria (Chen *et al.,* 2009). The cyt *b* is used to translocate the protein across the membranes for ATP synthesis and various cellular processes (Chen *et al.,* 2009). Since that it has been widely used as a primer, it has created its own status as a universal metric, in the sense that it can be manipulated to investigate phylogenetic relationship. The gene has a high variable and conserve within the population level but since that some part of the gene are more conserved than the others, which due to functional restrictions, the gene itself are prominent under strong evolutionary constraints (Meyer, 1994).

Meyer (1994) have even mentioned that regarded of the genes' functions and advantages, it also have disadvantages as well which will lead to giving information that is limited or less for understanding more on evolutionary process. Such disadvantages are bias in base compositional, variation between lineages, the third codon positions being very saturated, and variation in first and second codon positions are limited. Besides that, in addition to complete sequencings the whole cyt *b* gene, it took a lot of time and it is very laborious due to its size (1140 bp). Since that the fragment is in large size, powdered samples or processed products may not amplify the fragment which due to its highly degraded nature.

2.6 DNA Quantification

Simplest method in estimating DNA and RNA concentration would be the spectrophotometric quantification. To read a pure solution of DNA, it is quantitated by reading the absorbance at 260 nm. Through the ration of absorbance of 260 nm and 280 nm will give an estimation of the purity of the solution. Warburg and Christian (Warburg and Christian, 1942) was the first to describe the analysis of comparing the ration of the absorbance at A260 and A280 (260 nm and 280 nm). They came up with this in assessing the protein purity in the presence of nucleic acid contaminates.

Up until today, this method have been widely used to determine both of the presence of nucleic acid and protein as well as of its purity and yield. Nucleic acids absorb UV light with maximum as the four nucleotide components centered primarily at 260 nm. 1 cm light path gives the extinction coefficient of 20 for nucleotides. Through the advances technologies and great mind in invention, the calculation can be determined by the spectrophotometer machine itself. Evaluation of the spectrophotometric performance was done by Haque *et al.* (2003) to investigate the DNA quantification and fluorometric quantification method showed that the absorbance for the DNA was less biased than that fluorometric method that rely much to the multiple interacting component and more indirect measurement of DNA, in their investigation on human *BRRCAI* locus using PicoGreen assay (PG) and quantitative genomic PRC assay (QG).

Proteins can absorb UV light at 280 nm and 228 nm. The polypeptide chains which contains three amino acids with aromatic moieties, absorbs significantly at 280 nm, while as for histones or protamines give a little or no absorbance at 280 nm as they contain few or no aromatic residues. Peptide bonds absorb at 228 nm and are a more constant indicator of the presence of protein in a sample. Thus, absorbance readings measured both at 228 nm and at

280 nm provide a more accurate estimate of proteins or peptides that may be present in nucleic acid samples. For as carbohydrates, sulphydryls, phenolics and other aromatic compounds can also absorb at these wavelengths. Thus, from this information given, the presence of contaminant can be easily evaluated by comparing the ration of A260 reading to both the A280 and A228 measurements, rather than just determining by A260/A280 ratio. A260/ A280 ratio indicates the purity of DNA sample, which should be between 1.8 and 2.0 (Warburton *et al.*, 2005, Hochmeister, 1995). As for A260/A230, the value should be more than 2.0, since nucleic acids have minimum absorbances at 230 nm.

2.7 Previous Molecular Studies of Orang Utan

Warren *et al.* (2001) have done a research on speciation and intrasubspecific variation of Bornean Orang Utans from six different populations using mtDNA control region sequences and analysed whether there is any evidence of regional diversity. In their research data, they found that there are at least four subpopulations in Borneo;

(1) Southwest and Central Kalimantan,

(2) Northwest Kalimantan and Sarawak,

(3) Sabah, and

(4) East Kalimantan.

They suggested that genetic diversity between different groups is more and seems that the Bornean Orang Utans haven't face severe genetic bottleneck effect.

A research in comparing mtDNA sequences using NADH dehydrogenease subunit 3 and cytochrome B from the Bornean and Sumatran Orang Utans have been conducted by

Muir *et al.* (2000). They found out that the Orang Utan from Borneo form a tight cluster in their sequences, while those from Sumatra fall into three very distinct lineages. They suggested that the divergence in both of the genes within the Orang Utans in Sumatra must have resulted from an extended period of isolation of at least three groups and the presence of these groups is a relatively recent event.

In 1996, Zhi *et al.* (1996) investigate the genomic differentiation among the Orang Utan from Sumatra and Borneo. They suggested that there is no mutual understanding as to whether the Orang Utans' genetic differentiation is adequate enough to consider them as species, subspecies or population level taxonomic units, regardless they are separated subspecies based on their basic foundation of morphology, behaviour and cytogenetical characteristic.

Microsatellites and mtDNA sequences were used by Zhang *et al.* (2001) in their research to facilitate in their investigation on genetic divergence of both Bornean and Sumatran Orang Utans. Based on the data that they have collected, they predicted that both subspecies diverge around 2.3 ± 0.5 million years ago and both subspecies have not yet undergone a recent bottleneck.

Goossens *et al.* (2005) have conducted a research to investigate the patterns of genetic diversity and migration of the Orang Utan populations at the Lower Kinabatangan flood plain in Sabah, using fourteen microsatellite loci as genetic marker. They have collected hair and faecal samples from 200 wild Orang Utan in order to obtain their DNA. Based on the data that they have collected, they confirmed the river as a plays a role as natural barrier to gene flow. They also found out that the there is high probability of migration of the Orang Utan to the neighbouring area or across the river as they used to move relatively free.

3.0 Materials and Methods

3.1 Sample Collection

Samples consisting of hair of the Orang Utans were taken from Matang Wildlife Centre, Sarawak on the 13th of October 2010. Method used in obtaining the samples was pulling the hair from each individual, and then kept in different zip lock bag with labels according to their names given. All the samples were kept in the freezer below -20 °C in UNIMAS lab.

3.2 Cetyltrimethyl Ammonium Bromide (CTAB) Solutions

Cetyltrimethyl Ammonium Bromide (CTAB) solution was prepared before DNA extraction can be conducted. 40.9 g of Sodium Chloride (NaCl) was weighed and then was placed into 1000 mL beaker. Next, Tris-Base was then put into the beaker. This was followed by putting in CTAB and then ethlenediaminetetraacetic acid (EDTA), in an orderly manner. Then 400 mL of double distilled water ddH_2O was then added.

The beaker was placed on a hot plate and stirred until it became clear. The solution was left to cool. After that, it was then poured into 500 mL bottle and was proceed with wrapping the bottle with aluminium foil. 1000 µL of 2-mercaptoethanol-β-mercapto was added into the solution, followed by ddH_2O until the volume of the solution reached 500 mL.